

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-44191

(43) 公開日 平成9年(1997)2月14日

(51) Int.Cl.⁶
G 1 0 L 5/04
3/00

識別記号 序内整理番号

F I
G 10 L 5/04
3/00

技術表示箇所

審査請求 未請求 請求項の数9 OL (全 14 頁)

(21)出願番号 特願平8-39981
(22)出願日 平成8年(1996)2月27日
(31)優先権主張番号 特願平7-126491
(32)優先日 平7(1995)5月25日
(33)優先権主張国 日本(JP)

(71) 出願人 000001889
三洋電機株式会社
大阪府守口市京阪本通2丁目5番5号

(72) 発明者 西田 秀治
大阪府守口市京阪本通2丁目5番5号 三
洋電機株式会社内

(72) 発明者 平井 啓之
大阪府守口市京阪本通2丁目5番5号 三
洋電機株式会社内

(72) 発明者 宮武 正典
大阪府守口市京阪本通2丁目5番5号 三
洋電機株式会社内

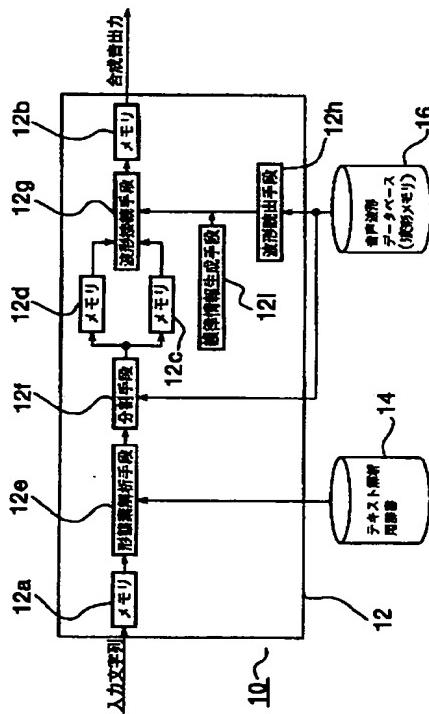
(74) 代理人 弁理士 岡田 敏

(54) 【発明の名称】 音声合成装置

(57) 【要約】

【課題】 従来の音声合成装置のデータベースに蓄積されている音声単位は、波形としてではなくケプスラム係数等のパラメータであるため、パラメータを音声波形に復元してから音声合成をしなければならず、音質の向上を妨げていた。

【解決手段】 入力文字列に対応する第1音韻記号列に含まれる複数の第1音韻記号部分列に対応する単位音声波形を合成音声出力する音声合成装置において、第1音韻記号列を所定の優先順位で複数の第1音韻記号部分列に分割する分割手段と、前記所定の優先順位で分割された第2音韻記号部分列を含む第2音韻記号列と前記第2音韻記号部分列に対応する単位音声波形を含む音声波形とを蓄積した波形メモリと、第1音韻記号部分列に対応する単位音声波形を前記波形メモリから読み出す波形読出手段と、および前記波形メモリから読み出された単位音声波形を接続して合成音声波形を生成する波形接続手段と、を備える。



【特許請求の範囲】

【請求項1】 入力文字列に対応する第1の音韻記号列に含まれる複数の第1の音韻記号部分列に対応する単位音声波形を合成して合成音声出力する音声合成装置において、第1の音韻記号列を所定の優先順位で複数の第1の音韻記号部分列に分割する分割手段と、

前記所定の優先順位で分割された第2の音韻記号部分列を含む第2の音韻記号列と前記第2の音韻記号部分列に対応する単位音声波形を含む音声波形とを蓄積した波形メモリと、第1の音韻記号部分列に対応する単位音声波形を前記波形メモリから読み出す波形読出手段と、および前記波形メモリから読み出された単位音声波形を接続して合成音声波形を生成する波形接続手段と、を備えることを特徴とする音声合成装置。

【請求項2】 前記所定の優先順位の要素として無音部、無声音部および有聲音部の少なくとも1つを含む、請求項1記載の音声合成装置。

【請求項3】 前記音声合成装置は前記入力文字列に対応する韻律情報を生成する韻律情報生成手段を備え、前記接続手段は前記単位音声波形を前記韻律情報に従って波形整形する波形整形手段を含む、請求項2記載の音声合成装置。

【請求項4】 前記韻律情報はピッチパターンおよびパワーパターンの少なくとも一方を含む、請求項3記載の音声合成装置。

【請求項5】 前記有聲音部で分割された単位音声波形を接続するとき、前記接続手段はそれぞれの単位音声波形に含まれる1ピッチ分の波形に適当な重み付けをして加算する加算手段を含む、請求項2ないし4のいずれかに記載の音声合成装置。

【請求項6】 前記波形メモリから読み出された単位音声波形を接続する時、その接続歪みが小さくなるように、前記第1の音韻記号列を分割する音韻分割点前後の音韻の種類が一致するような単位音声波形を選択することを特徴とする請求項2記載の音声合成装置。

【請求項7】 前記波形メモリから読み出された単位音声波形を接続する時、前記第1の音韻記号列を分割する音韻分割点数が最小になるような単位音声波形を選択することを特徴とする請求項2記載の音声合成装置。

【請求項8】 前記波形メモリから読み出された単位音声波形を接続する時、その接続歪みが小さくなるように、その接続点における前後の波形のピッチが所定量以上異なるように前記単位音声波形を選択することを特徴とする請求項2ないし4のいずれかに記載の音声合成装置。

【請求項9】 前記波形メモリから読み出された単位音声波形を接続する時、その接続歪みが小さくなるように、前記第1の音韻記号列を分割する音韻分割点の優先順位、音韻分割点前後の音韻の種類の一致度、音韻分割点数、及び接続点前後の波形のピッチの差を夫々量子化

し、その量子化された夫々の値に所定の重みを付加した評価関数により前記単位音声波形を選択することを特徴とする請求項2ないし4のいずれかに記載の音声合成装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は音声合成装置に関し、特にたとえば音声案内、音声応答および音声読み上げ等に用いられ、入力文字列に対応する音韻記号列に従って音声波形を合成して出力する、音声合成装置に関する。

【0002】

【従来の技術】従来のこの種の音声合成装置が、平成4年12月22日に出願公開された特開平4-369693号公報に開示されている。これは、音声合成を行う音韻系列に対応するかつデータベースに含まれる音声単位の中から、平均的な音響特性を有する音声単位を優先的に選択して、それぞれを接続するものである。

【0003】また、音声波形を編集して音声合成する方法としては、発表論文「波形辞書を用いた規則合成法」(電子情報通信学会音声研資SP88-9, pp65-pp72, (1988年), 発表者: 広川智久)が周知となっている。これには、類似する音声波形から代表音声波形を抽出し、その代表音声波形を編集して音声合成をする方式や、繁茂に出現する音声波形をデータベース化し、発生頻度の高い音声波形単位を編集して音声合成をする方式が開示されている。

【0004】

【発明が解決しようとする課題】しかし、特開平4-369693号公報の音声合成装置では、著しく異なった音響特性を有する音声単位が選択されることによる合成音の音質劣化はなくなるものの、相変わらず音質は優れているとはいえないかった。

【0005】また、それぞれの音声単位は、波形としてではなくケプスラム係数等のパラメータに変換してデータベースに蓄積されているため、パラメータを音声波形に復元してから音声合成をしなければならず、これも音質の向上を妨げる原因となっていた。

【0006】更に、「波形辞書を用いた規則合成法」による音声合成方式では、いずれの方式も音節などを単位としているため、合成音の音質が優れているとはいえないかった。

【0007】それゆえに、この発明の主たる目的は、優れた音質の合成音を出力することができる、音声合成装置を提供することである。

【0008】

【課題を解決するための手段】本発明は、入力文字列に対応する第1の音韻記号列に含まれる複数の第1の音韻記号部分列に対応する単位音声波形を合成して合成音声出力する音声合成装置において、第1の音韻記号列を所

定の優先順位で複数の第1の音韻記号部分列に分割する分割手段、所定の優先順位で分割された第2の音韻記号部分列を含む第2の音韻記号列と第2の音韻記号部分列に対応する単位音声波形を含む音声波形とを蓄積した波形メモリ、第1の音韻記号部分列に対応する単位音声波形を波形メモリから読み出す波形読出手段、および波形メモリから読み出された単位音声波形を接続して合成音声波形を生成する波形接続手段を備えることを特徴とする音声合成装置である。

【0009】

【発明の実施の形態】本発明の実施の形態を図1乃至図9に基づいて説明する。

【0010】図1を参照して、この実施例のテキスト音声合成装置10は、マイコン12を含む。マイコン12は、テキスト文章データからなる入力文字列を受け、まずテキスト解析用辞書14を用いて、形態素解析手段12eにて入力文字列の形態素解析を行う。而して、分割手段12fにて入力文字列を音韻記号部分列に分割して、その入力文字列を分割点が設定された音韻記号列に変換するとともに、この入力文字列のピッチパターンおよびパワーパターンを生成する。

【0011】次に、波形接続手段12gは、波形読出手段12hにて音声波形データベース16から読み出された単位音声波形をピッチパターンおよびパワーパターンに基づいて、整形および接続編集し、これによって生成された合成音を出力する。

【0012】音声波形データベース16には、音声波形と各々の音声波形に対する音韻ラベル情報とが登録されている。音韻ラベル情報は音韻記号列および記号列番号を含む。この実施例では、以下の記号列番号および音韻記号列が音声波形データベース16に登録される。なお、音韻記号列に含まれる“-”は20 msec以上の無音区間を表している。

【0013】(1)-n i N g e N -t o w a -k a N g
a e r u a s i -d e a r u - (人間とは、考える葦である)

(2)-k o r e w a -b i m y o -d e a r u - (これは、微妙である)

(3)-f u -k u z a -t u - (複雑)

(4)-y u n y u s i -t e f u -k a s a s e - (輸入してふ化させ)

(5)-t e o -t u n a i -d e u -t a u - (手をつないで歌う)

(6)-h i -t o i -k i r e n i - (人いきれに)

(7)-m i n i m o a r u - (ミニもある)

(8)-g e N s a -k u s o n o m o n o - (原作そのもの)

(9)-s y o -k o d e a r u - (証拠である)

ここで、入力文字列に対応する音韻文字列、パワーパターンおよびピッチパターンを生成するためのアルゴリズム

ムを図2に示す。

【0014】マイコン12はまず、ステップS1で、入力文字列を1文単位でメモリ12aに書き込む。次に、ステップS3で形態素解析手段12eにて入力文字列の形態素解析をする。すなわち、テキスト解析用辞書14には、単語の表記とそれに対する音韻記号列(読み)やアクセント、品詞などの情報が蓄えられているため、これを用いて入力文字列がどのような単語から成り立っているかを解析する。続いて、ステップS5で解析結果に基づいて入力文字列の音韻記号列を生成する。

【0015】その後、ステップS7でテキスト解析用辞書14を用いて入力文字列のポーズ(PAUSE)情報を解析し、この解析結果からステップS9で入力文字列のパワーパターンを生成する。

【0016】さらに、ステップS11で、テキスト解析用辞書14を用いて入力文字列のアクセント情報を解析し、この解析結果からステップS13で入力文字列のピッチパターンを生成する。

【0017】ここで、「人間とは、微妙で複雑な生き物である」という文字列が入力された場合のパワーパターンおよびピッチパターンを図3に実線で示す。パワーパターンは、それぞれの音韻記号に対応する単位音声波形の振幅倍率($0 \leq p_w \leq 2$)を時間関数 $p_w(t)$ で表したものである。なお、音韻継続時間は日本語モーラ時間単位で換算されている。

【0018】一方、音声波形データベース16に登録された音声波形に含まれるそれぞれの単位音声波形は、予め平均振幅が基準振幅“1”となるように正規化されている。すなわち、音声波形は20 kHzサンプリングの16bit PCMデータ(量子化値は、-32768-32767の範囲をとり得る)として登録されており、これに含まれる単位音声波形の平均振幅が16384になるように正規化されている。

【0019】図3からわかるように、入力文字列に対応する音韻記号列に含まれる/n i N g e n/の頭の部分、および/i k i m o n o/の全体の振幅倍率を強調のため上げているが、その他の部分は振幅倍率1となっており、この部分については音声波形データベース16に登録されている単位音声波形の振幅を変えずにそのまま接続編集することができる。参考までに、波形整形後の振幅変化を点線で示す。また、ピッチパターンは音声波形の周波数($50 \leq p_t \leq 300$ [Hz])を時間関数 $p_t(t)$ で表したものであり、周知のモデル(藤崎モデル:比企静雄編、「音声情報処理」東京大学出版会、1973)により算出されている。

【0020】次に、入力文字列に対応する音韻記号列、パワーパターンおよびピッチパターンに基づいて出力音声を生成するアルゴリズムを図4に示す。

【0021】分割手段12fは、まずステップS15で入力文字列に対応する音韻記号列の分割点を決定し、こ

の音韻記号列を複数の音韻記号部分列に分割する。

【0022】次に、ステップS17で波形接続手段12gは部分列ナンバーnを“1”に設定し、さらにステップS19で、波形読出手段12hはn番目の音韻記号部分列に対応する単位音声波形およびラベル情報を音声波形データベース16より抽出する。

【0023】続いて、ステップS21で、波形接続手段12gは入力文字列に対応するパワーパターンに一致するように単位音声波形の音韻継続時間長およびゲインを波形整形によって修正する。

【0024】その後、ステップS23で、波形接続手段12g中の波形整形手段は、入力文字列に対応するピッチパターンに対応するように単位音声波形のピッチを波形整形によって修正する。

【0025】続いて、ステップS25で波形を接続し、接続した合成音声波形をステップS27でメモリ12bに記憶する。その後ステップS29で部分列ナンバーnをインクリメントし、ステップS31でn番目の単位音声波形が存在するかどうか判断する。ここで“YES”であればステップS19に戻るが、“NO”であればステップS33で合成音声波形をアナログ音声波形に変換して出力する。

【0026】ステップS15は、具体的には図5に示すアルゴリズムによって表される。マイコン12は、まずステップS1501で音韻記号列を解析し、無音部分割点を優先順位第1位の分割点として検索する。

【0027】すなわち、音響学的解析によってその音韻記号列を音声に変換して出力したとき20 msec以上の無音が継続する箇所を無音部分割点とし、この点を検索する。

【0028】具体的には、文脈解析によって検出できるボーズ発生箇所の他に、日本語発声学的な解析によって、促音“Q”や子音“p”、“t”、“k”、“d”、“b”、“z”等の直前にクロージャと呼ばれる20～70 msec程度の無音区間箇所があることがわかつており、これらを音響学的解析により判定する。

【0029】したがって、この実施例の音韻記号列については、ステップS1501で以下の無音部分割点が検索される。ここでは、無音部分割点を／で表し、無音部を“-”で表している。

【0030】/-n i N g e N -/-t o w a -/-b i m y o -/-d e f u -/-k u z a -/-t u n a i -/-k i m o n o d e a r u -/

次にステップS1503で、“s”、“sh”、“h”、“hy”、“f”等の無聲音部分（無声子音部分）を優先順位第2位の分割点とし、これを検索する。したがって、この実施例の音韻記号列では以下のような分割点が選出される。区別のために無聲音部分割点を//で表す。

【0031】/-n i N g e N -/-t o w a -/-b

i m y o -/-d e f //f u -/-k u z a -/-t u n a i -/-k i m o n o d e a r u -/さらに、ステップS1505では、“a”、“i”、“u”、“e”、“o”等の有声音部分（母音部分）を優先順位第3位の分割点とし、これを検索する。したがって、この実施例の音韻記号列では、以下のような分割点が追加される。区別のため、有声音部分割点を///で表す。

【0032】/-n i /// i N g e /// e N -/-t o //ow a -/-b i /// i m y o -/-d e /// e f //f u //u -/-k u //u z a //a -/-t u //u n a //a i -/-k i //i m o //o n o //o d e //e a //a r u -/

したがって、この音韻記号列は、最大23個の音韻記号部分列に分割可能となる。なお、音声波形データベース16に登録されている音韻記号列も無音部分割点、無聲音部分割点および有声音部分割点を持ち、音声波形はこれらの分割点に対応する波形分割点を持つ。

【0033】分割手段12fは以下のステップS150207～S1529で、優先順位を考慮して、なるべく音韻記号部分列の数が少なくなるように、最終的な分割点を決定していく。ステップS1507では、分割手段12fは無音部分割点によって分割される音韻記号部分列の1つをメモリ12cに書き込む。本発明の実施の形態では、まず最初に入力音韻記号列の頭から次の優先順位第1位までの分割点までの以下の音韻記号部分列がメモリ12cに書き込まれる。

【0034】/-n i N g e N -/

次に、ステップS1509で、分割手段12fは音声波形データベース16のラベル情報にステップS1507で書き込んだ音韻記号部分列に対応する音韻記号部分列が含まれていないかどうか判断する。そして、分割手段12fは音韻記号部分列が存在すれば、ステップS1511で分割点を決定する。

【0035】すなわち、分割手段12fは部分列ナンバーn及び、その音韻記号部分列を含む音声波形データベース16内の音韻記号列の記号列番号1（本実施の形態では(1)～(9)）とその音韻記号部分列に対応する音声波形の開始点sと終了点eの情報をメモリ12dに書き込む。その後、ステップS1507に戻り、次の音韻記号部分列に対しての処理をする。

【0036】この実施例では、/-n i N g e N -/は音声波形データベース16の中の1番目の音韻記号列に存在するので、分割点/-n i N g e N -/が決定され、メモリ12dに/-n i N g e N -/の存在する記号列番号1と対応する音声波形の波形開始点及び終了点が記憶される。

【0037】その後ステップS1507に戻り、同様に次の音韻記号部分列/-t o w a -/をメモリ12cに上書きし、ステップS1509で音声波形データベース

16に含まれる音韻記号部分列を検索する。ここで、／-t o w a-／は音声波形データベース16の中の1番目の音韻記号列に存在するので、ステップS1511で分割点／-t o w a-／が決定される。同様に次の音韻記号部分列／-b i m y o-／も、音声波形データベース16の中の2番目の音韻記号列に存在するので、分割点／-b i m y o-／が決定される。

【0038】このようにして、／-n i N g e N-／-t o w a-／-b i m y o-／の分割点が決定される。

【0039】ステップS1509で“NO”と判断されると、分割手段12fはステップS1513で、優先順位第2位の分割点、すなわち無声音部分割点によって分割されるなるべく長い音韻記号部分列を抽出し、これをメモリ12cに書き込む。

【0040】次に、ステップS1515で、ステップS1509と同様に、メモリ12cに記憶された音韻記号部分列が音声波形データベース16に存在するかどうか判断する。

【0041】ここで“YES”であれば、ステップS1517でステップS1511と同様にして、分割手段12fは無声音部による分割点を決定し、ステップS1507に戻るが、“NO”であれば、ステップS1519で無声音部分割点によってこれ以上の分割が可能であるかどうか判断する。そして、“YES”であればステップS1513に戻り、“NO”であればステップS1521に移行する。

【0042】本発明の実施の形態では、次の音韻記号部分列／-d e f u-／は音声波形データベース16に存在しないので、ステップS1513では／-d e f //がメモリ12cに上書きされ、次にステップS1515でこの音韻記号部分列が音声波形データベース16に存在するかどうか判断する。

【0043】しかし、／-d e f //は音声波形データベース16に存在しない。この実施例では、／-d e f //の音韻記号列は音声波形データベース16に存在せず、これ以上優先順位第2位での分割点による分割が不可能なため、ステップS1519を経てステップS1521に移行する。

【0044】ステップS1521では、優先順位第3位の分割点、すなわち有声音部分割点で分割されるなるべく長い音韻記号部分列を抽出し、これをメモリ12cに書き込む。次にステップS1523で、ステップS1509およびS1515と同様に、この音韻記号部分列が音声波形データベース16に含まれているかどうか判断する。そして、含まれていれば、ステップS1525でステップS1511およびS1517と同様にして有声音部分割点を決定し、ステップS1507に戻るが、含まれていなければ、ステップS1527で有声音部分割点によってこれ以上の分割が可能であるかどうか判断する。

【0045】そして、可能であればステップS1521に戻り、不可能であればステップS1529でエラー処理をしてステップS1507に戻る。すなわち、ステップS1529では、ステップS1527で最終的にこれ以上分割できないと判定された音韻記号部分列を無視して、検索すべき音韻記号部分列を更新し、ステップS1507に移る。

【0046】本発明の実施の形態では、ステップS1521において以下の如く判定する。ここでは、／-d e // / / / が音声波形データベース16の1、2、及び5番目の音韻記号列に存在する。しかし、音韻記号部分列は1番目の音韻記号列から昇べき順に検索され、見つかった時点での検索は中止されるため、ステップS1521において／-d e // / / / は2番目の音韻記号列から選ばれ、ステップS1525で対応する分割点が決定される。もし、ステップS1527でこれ以上分割不可能と判定されれば、ステップS1529でエラー処理をし、ステップS1507に戻る。

【0047】以上の処理により、最終的には以下のように分割点が決定される。

【0048】／-n i N g e N-／-t o w a-／-b i m y o-／-d e // / / / e f // / f u-／-k u z a-／-t u n a i-／-k i // / i m o // / o n o // / o d e a r u-／

他方、分割手段12fは分割点の決定に関して、次の手法を採用することも可能である。

【0049】本手法では、入力音韻記号列の分割点の組み合わせからできるすべての音韻記号部分列について以下に示す「評価関数score」により決定される評価点を算出し、各々の音韻記号部分列に対応する評価点の累積が最小になる組み合わせより分割点を決定する。

【0050】ここで、評価関数scoreは、分割点の優先順位により決定される値type、分割点前後の音韻の種類により決定される値link、分割された音韻長により決定される値len、及び分割点に対応する波形接続点における理論値とのピッチ周期の差により決定される値f0のそれぞれの数値にw1～w4の重みをかけて足し合わせた値とする。w1～w4の重みは、それぞれ0～1までの実数定数である。

【0051】評価関数： $score = w1 * type + w2 * link + w3 * len + w4 * f0$

但し、type = 0（分割点が前記優先順位第1位である場合）

type = 1（分割点が前記優先順位第2位である場合）

type = 3（分割点が前記優先順位第3位である場合）

type = 9（それ以外の場合）

link = 0（分割点前後の音韻の種類が一致する場合）

link = 9（それ以外の場合）

len = -（分割点で区切られた音韻記号部分列の音韻記号数）

$f_0 = |\log(\text{実波形のピッチ周期}) - \log(\text{理論ピッチ周期})|$

である。

【0052】以下、入力文字列／－a m e n o t a m e k a－／（雨のためか）について、分割点の決定方法について述べる。説明の簡略のため本実施例では、 $w_1=1$, $w_2=1$, $w_3=1$, $w_4=0$ とした。

【0053】音韻記号部分列の組み合わせは、図7に示す木検索により行う。

【0054】同図中、選択された音韻記号部分列（この音韻記号部分列は音声波形データベース16のラベル情報に存在し、かつ、すべての音韻分割点前後の音韻が一致するものが選択されたとした）の下側にscore値が示されている。説明のため、各音韻記号部分列の選択された状態を便宜上ノード0からノード8と呼ぶ。

【0055】まず、はじめにノード0において／－／（無音）で始まり、／－a m e n o...／と続く音韻記号部分列を音声波形データベース16のラベル情報より検索し、その中で最もscore値が小さい音韻記号部分列から所定の数m個（本実施の形態では、2個とする）選択し、下位にm個のノードを作成する。図7では、ノード1／－a m e n o－/とノード4／－a m e n o－t a m//が選択された。／－a m e n o－/のscore値は、

type = 9：優先順位外の分割点で終わっている。

【0056】link = 0：後続の音韻記号がtで一致している。

【0057】len = -10

score = $9 + 0 - 10 = -1$

／－a m e n o－t a m/のscore値は、

type = 0：優先順位第1位の分割点で終わっている。

【0058】link = 0：後続の音韻記号がmで一致している。

【0059】len = -7

score = $0 + 0 - 7 = -7$

である。ここで、ノード1およびノード4を音韻分割部分列候補とする。従って、それぞれのノードでの累計score値は、それぞれ

ノード1での累計score = -1

ノード4での累計score = -7

となる。分割毎に累計scoreの小さいものからm個の音韻部分列の探索系列を残すため実施例ではノード1およびノード4の音韻部分列は候補として残る。従って、次の探索として、ノード2、3、5及び6が候補となり、ノード2での累計score = -3

ノード3での累計score = -7

ノード5での累計score = -7

ノード6での累計score = -9

この場合、ノード3、5が同点であるが、同点の場合は

そのノードでのscore値の小さいほうを優先し、結果としてノード3、6が候補として残る。

【0060】ここでノード3は分割が終了したので、ノード3での累計scoreは、常に候補として残る。ノード2および5からの探索はこれ以上行わない。同様に分割を繰り返し最終的に残ったノードは、図7より、ノード3およびノード8となり、それぞれの累積score値は、ノード3での累計score = -7
ノード8での累計score = -14

10 であるから、scoreが小さいほうのノード8までの検索による音韻分割が最適となり、実際の分割は、／－a m e n o－/－t a m e / e -k a -/に決定される。

【0061】次に、図4のステップS21およびS23の波形整形に関する具体的な処理について説明する。

【0062】マイコン12は、ステップS1511, S1517およびS1525で決定された分割点に従って、波形読出手段12hは音声波形データベース16に登録されている単位音声波形を抽出し、波形接続手段12gはパワーパターンおよびピッチパターンによってそ

20 の単位音声波形を整形する。

【0063】単位音声波形の振幅および音韻継続時間長はパワーパターンpw(t)に従って、また、音程(ピッチ)はピッチパターンpt(t)に従って整形される。単位音声波形の振幅制御においては、時刻tにおけるpw(t)を基準として振幅ゲインを調整する。音韻継続時間長および音程制御は、PSOLA法(F.Charpentier他、「Pitch-Synchronous Waveform Processing Techniques for Text-to-speech Synthesis Using Diphones」Proc. Eurospeech '89を用いる。

30 【0064】たとえば、音韻記号列／－d e///に対応する単位音声波形の抽出と波形整形について述べる。音声波形の分割点に関してより詳しく説明すると、無音部分割点は無音区間の開始点あるいは終了点に設けられ、無声音部分割点は無声音区間のほぼ中心に設けられ、そして、有声音部分割点は周期性が安定している母音中心部等の1ピッチ波形の開始点と終了点とにおける右下がりのゼロクロスポイントに設けられる。

【0065】そして、これらの波形分割点がマーキングされるとともに、そのマーキングが各々の音声波形の先頭サンプルから何サンプル目にあるかが、上述のステップS1511, S1517およびS1525でメモリ12dに書き込まれる。

【0066】したがって、単位音声波形／－d e///の波形終端は、音韻記号“e”に相当する音声波形のうち周波数の安定している1ピッチの終了点であり、この波形終端までの波形が単位音声波形として抽出される。抽出された単位音声波形は、前述の方法により波形整形される。

【0067】このうち、音程制御にあたっては、接続箇所である1ピッチ波形の終了点を基準時とし、その基準

時の基準ピッチ周期Tをピッチパターンに基づいて算出する。

【0068】ここで、基準ピッチとは入力文字列に対応する音韻記号列の当該基準時におけるピッチ周期である。そして、算出された基準ピッチ周期Tにマーキングされた1ピッチ波形のピッチ周期が一致するように単位音声波形全体をPSOLA法で整形する。

【0069】このとき、次の単位音声波形の接続のために、波形整形の後に接続された単位音声波形の波形分割点は、ステップS23で合成音声波形の始めからのサンプル数に換算して記憶保持される。

【0070】また、音韻継続時間長の制御にあたっては、パワーパターンに合わせて、無音区間の補間・削除またはPSOLA法によって、ピッチが安定している母音区間における1ピッチ波形を補間または間引きする。振幅の調整にあたっては、音声単位波形毎にパワーパターンより導かれる振幅倍率を乗じる。

【0071】統いて、図4のステップS25における波形接続処理について説明する。有聲音区間の接続に関しては、互いに接続する単位音声波形の特定の1ピッチがマーキングされているため、この期間Tにおいて、それぞれの波形に重みをかけて足し合わせることによって、両波形を接続する。

【0072】すなわち、先行する単位音声波形の特定の1ピッチに対しては、1から0へ直線的に変化する重み(窓関数 $W_1 = 1 - i / (T - 1)$ ($i = 0, 1, \dots, T - 1$))をかけ、後行する単位音声波形の特定の1ピッチに対しては、0から1へ直線的に変化する重み(窓関数 $W_2 = i / (T - 1)$)をかけて、両波形を足し合わせる。また、無音区間の波形接続および無声音区間の波形接続に関しては、互いに接続する単位音声波形のマーキングされた分割点で単純に接続する。

【0073】たとえば、 $/ - d e // /$ と $/ / / e f //$ の接続に関して、図8を用いて説明する。図8(a)は、音声波形データベース16に登録されている $/ - d e // /$ (2番目に存在)の音韻記号“e”に相当する波形整形済みの音声波形の一部分であり、図8(b)は、音声波形データベース16に登録されている $/ / / e f //$ (4番目に存在)の音韻記号“e”に相当する波形整形済みの音声波形の一部分である。この2つの波形を接続して、 $/ - d e f //$ に相当する合成波形である図8(c)を得る。

【0074】この場合は有聲音区間の波形接続であるので、互いに接続する図8(a)、(b)各々の単位音声波形の1ピッチ分がA、BおよびC、Dでマーキングされており、図の破線で示すように、各々の波形に重みを付けて足し合わせることにより両者が接続される。この

とき、音声波形の図8(a)、(b)の接続箇所AからB、CからDの時間長は、前述の基準ピッチ周期Tになるよう波形整形されている。

【0075】次に、たとえば、 $/ / / e f //$ と $/ / f u //$ の接続に関して、図9を用いて説明する。

【0076】図9(d)は、音声波形データベース16に登録されている $/ / / e f //$ (4番目に存在)の音韻記号“f”に相当する波形整形済みの音声波形の一部分であり、図9(e)は、音声波形データベース16に登録

10 されている $/ / f u //$ (3番目に存在)の音韻記号“f”に相当する波形整形済みの音声波形の一部分である。この2つの波形を接続して、 $- d e f u - /$ の“f”に相当する合成波形である図9(f)を得る。この場合は、無声音区間の波形接続であるので、互いに接続する図9(d)、(e)各々の単位音声波形のマーキングされた波形分割点E、Fで単純に接続する。

【0077】

【発明の効果】この発明によれば、所定の優先順位で分割された第1の音韻記号部分列に対応する最適な単位音声波形が読み出手段によって波形メモリから読み出され、波形接続手段によって接続されるため、音質が優れた合成音声波形を出力することができる。

【図面の簡単な説明】

【図1】本発明の一実施の形態を示すブロック図である。

【図2】本発明の動作の一部を示すフロー図である。

【図3】音韻記号列、パワーパターンおよびピッチパターンを示す図解図である。

【図4】本発明の動作の一部を示すフロー図である。

【図5】本発明の動作の一部を示すフロー図である。

【図6】本発明の動作の一部を示すフロー図である。

【図7】本発明の動作の一部を示す木構造図である。

【図8】(a)および(b)は単位音声波形の一部を示す波形図であり、(c)は合成された音声波形の一部を示す波形図である。

【図9】(d)および(e)は単位音声波形の一部を示す波形図であり、(f)は合成された音声波形の一部を示す波形図である。

【符号の説明】

40 10 …テキスト音声合成装置

12 e …形態素解析手段

12 f …分割手段

12 g …波形接続手段

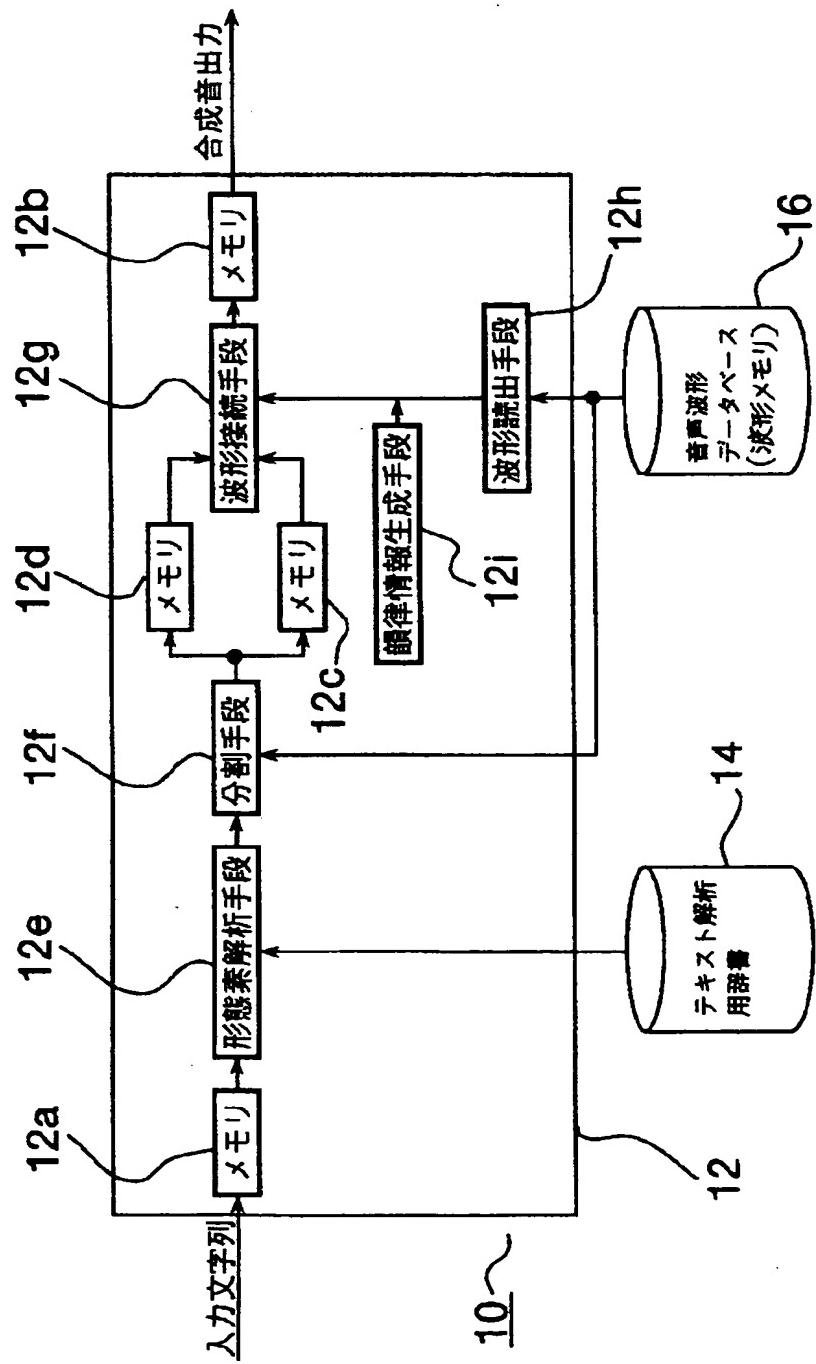
12 h …波形読み出手段

12 i …韻律情報生成手段

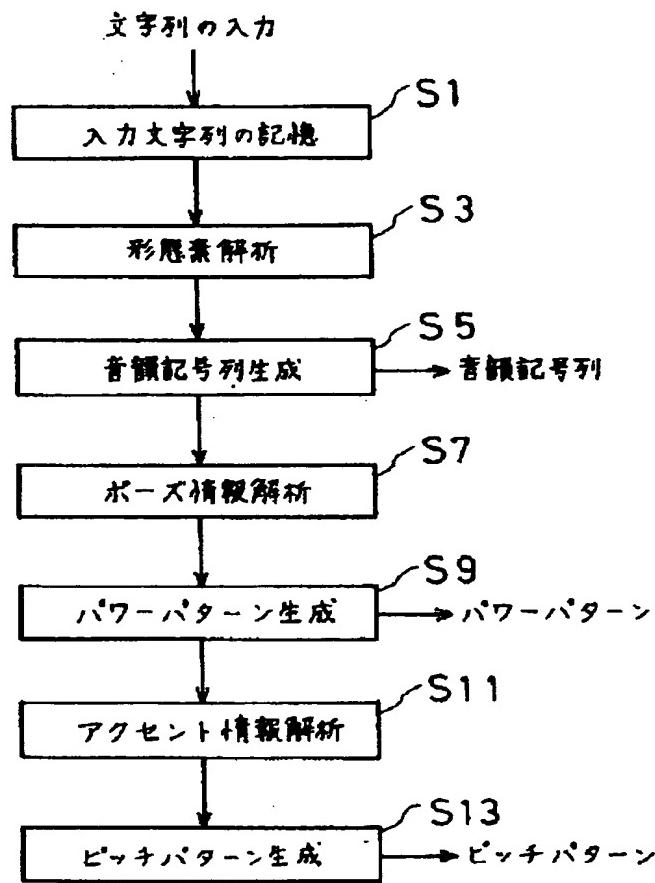
14 …テキスト解析用辞書

16 …音声波形データベース

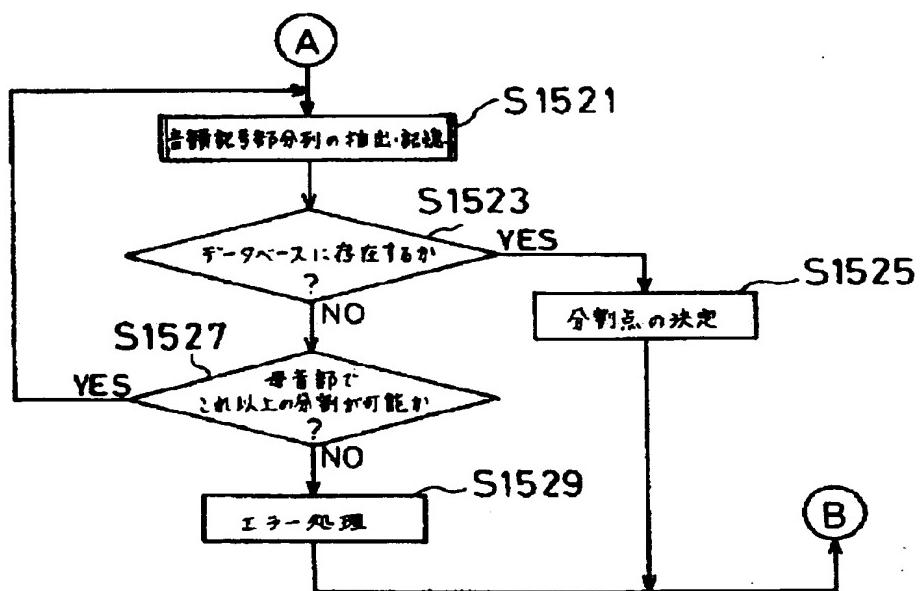
【図1】



【図2】

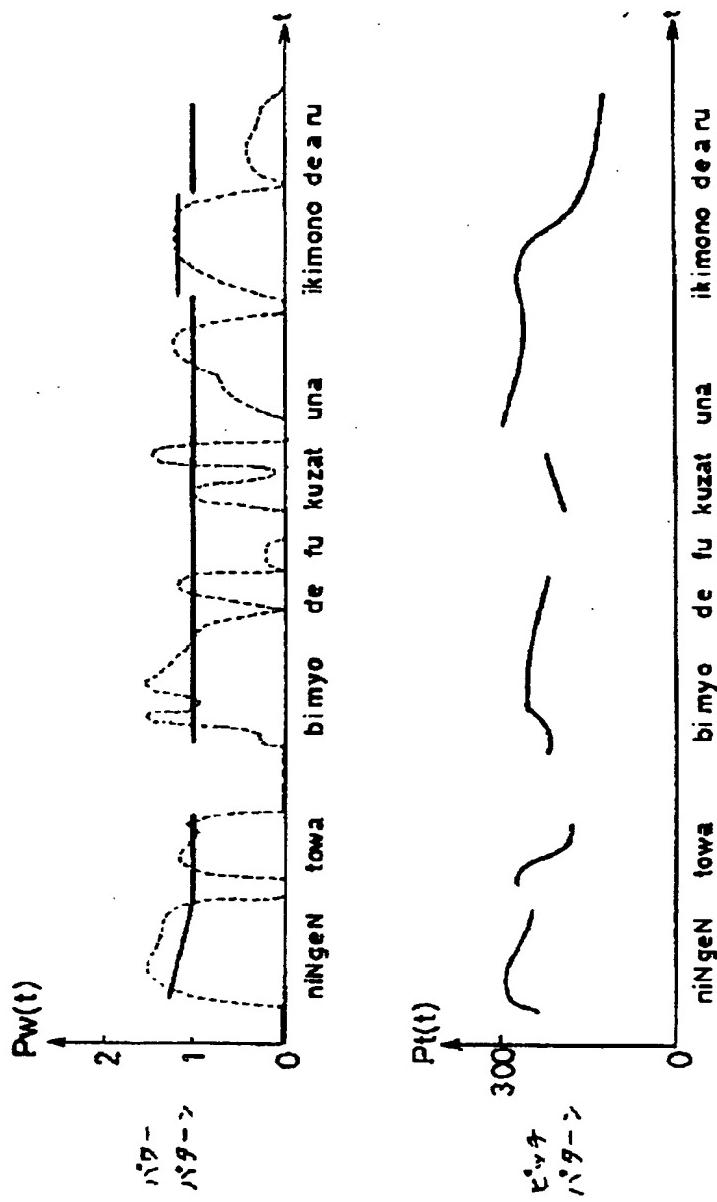


【図6】

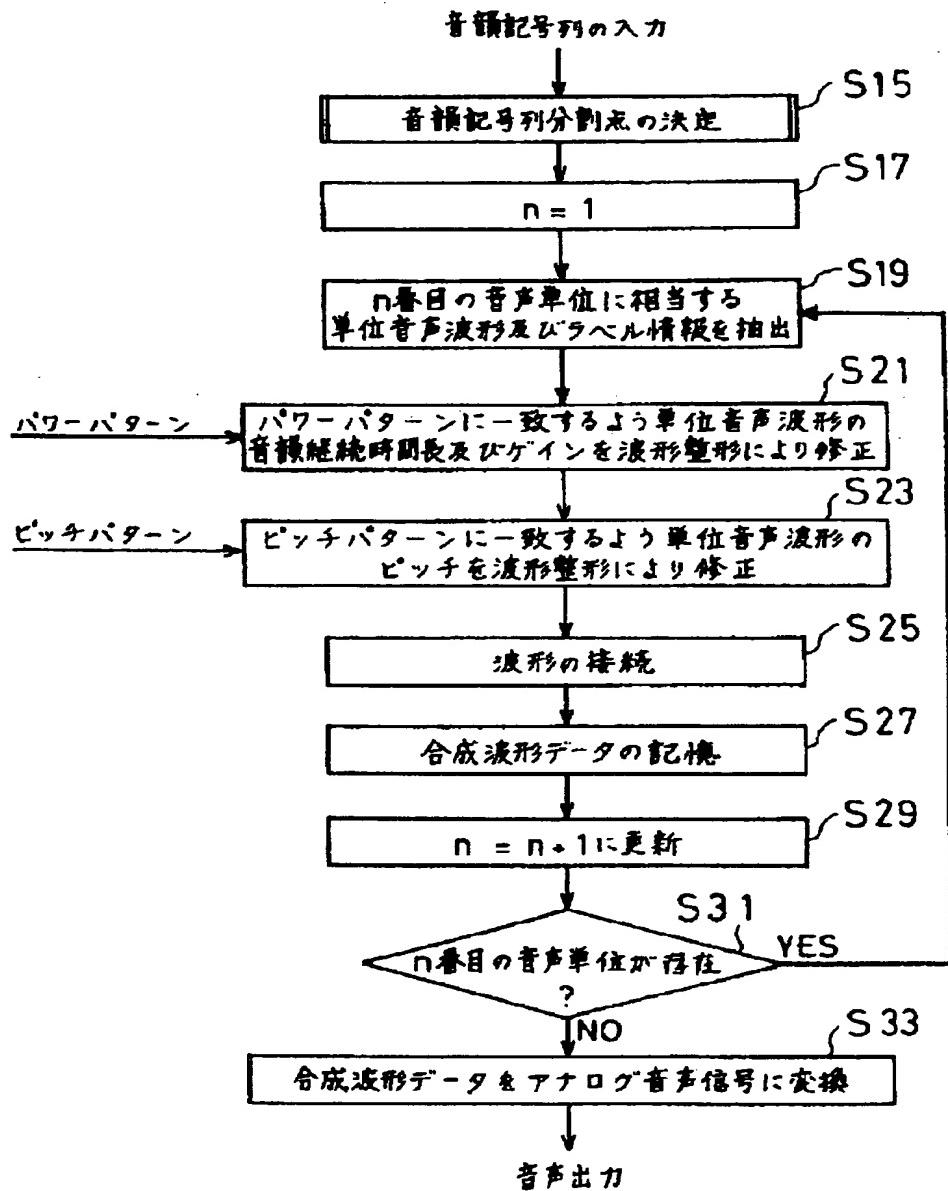


【図3】

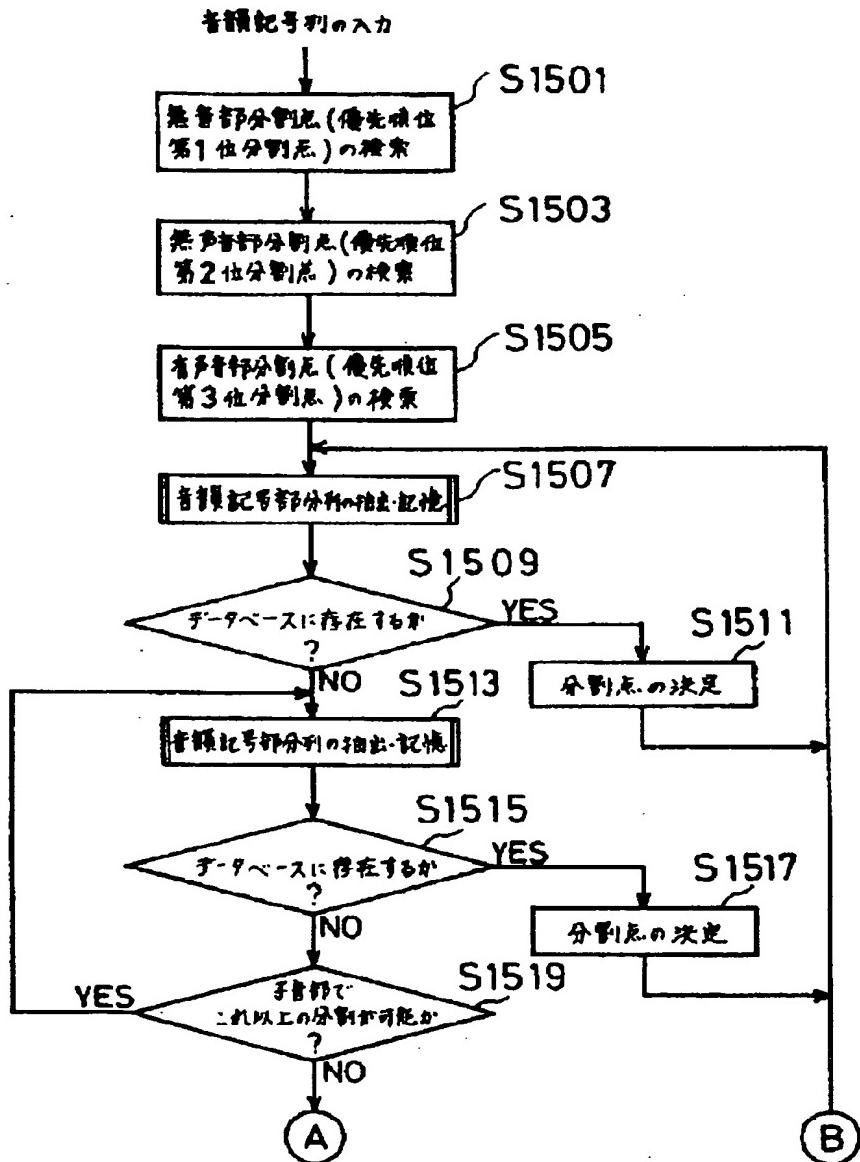
音記号： niNgen towa bimyo de fu kuzat una ikimono de aru.



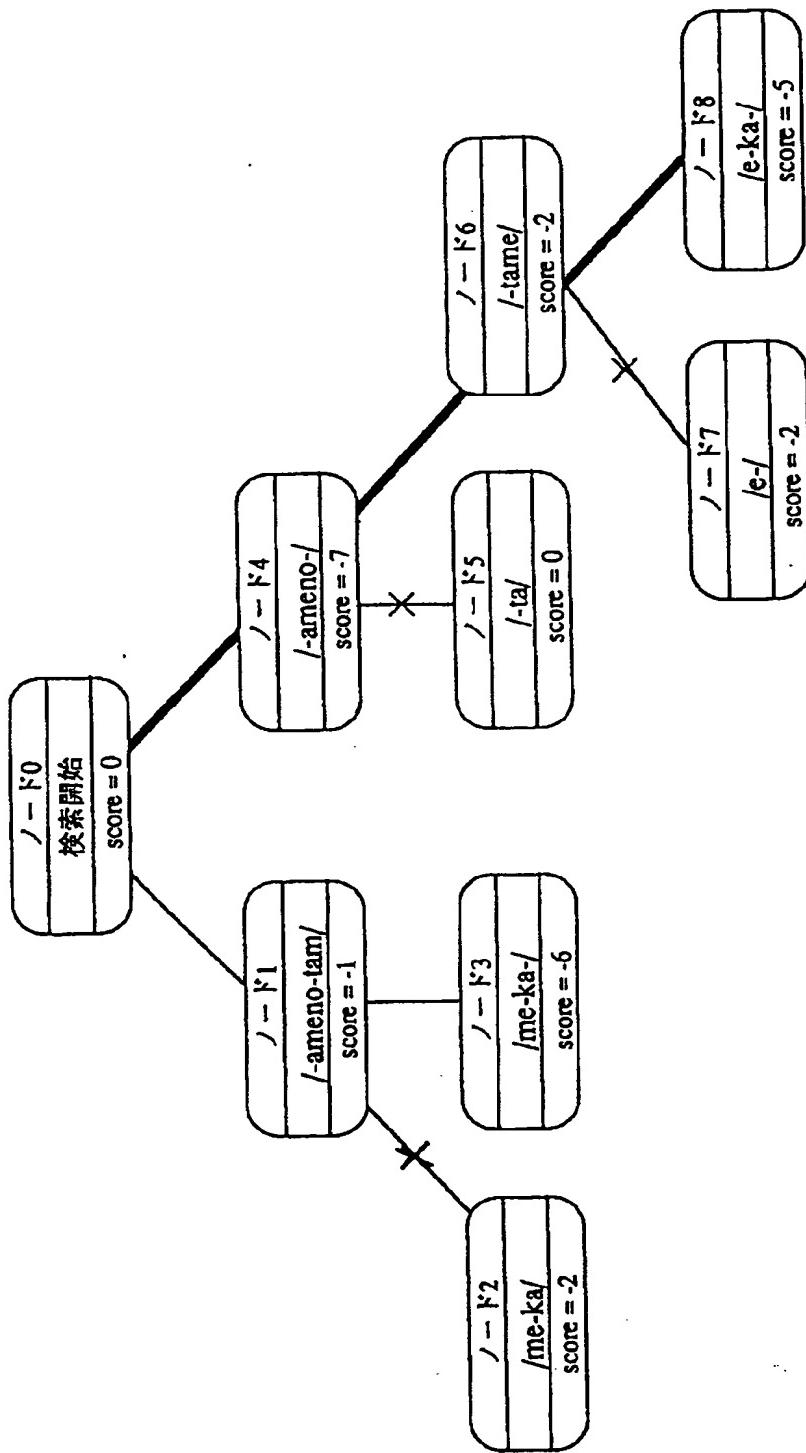
【図4】



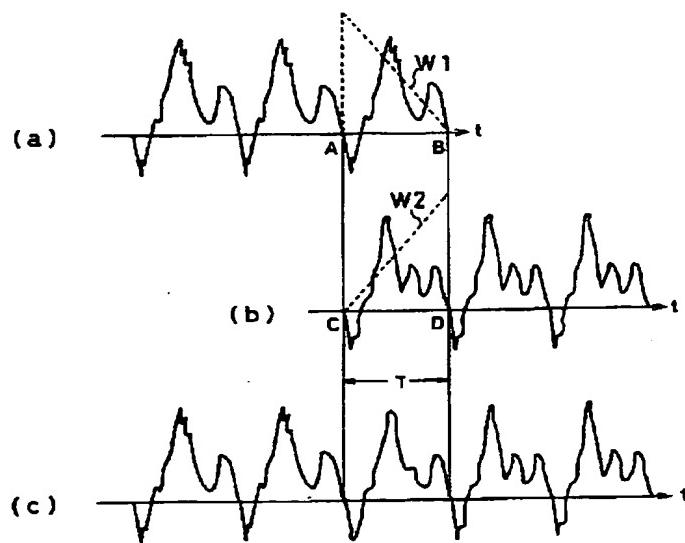
【図5】



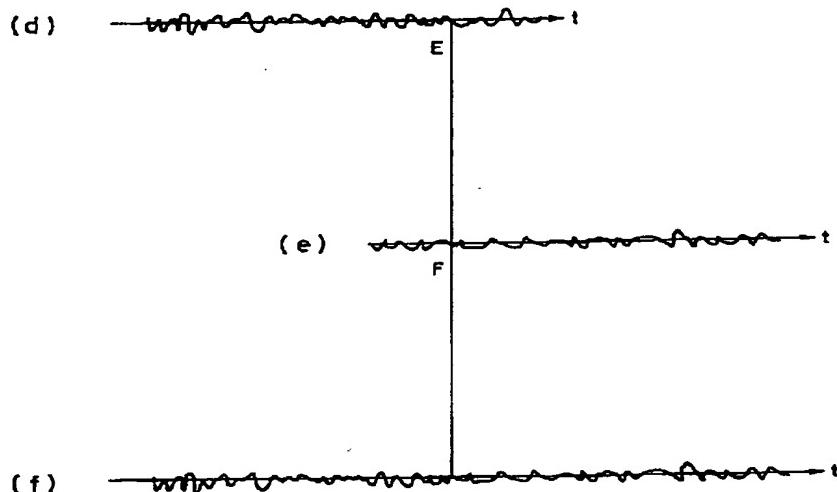
【図7】



【図8】



【図9】



フロントページの続き

(72)発明者 大西 宏樹
 大阪府守口市京阪本通2丁目5番5号 三
 洋電機株式会社内